Bayesian Estimation of Key Population Sizes Using Multiple Data Sources

by

Andrew Pita

A thesis submitted to The Johns Hopkins University

in conformity with the requirements for the degree of

Master of Science

Baltimore, Maryland

April, 2019

Abstract

Monitoring of key populations with high HIV prevalence and high risk of infection is an essential aspect of HIV prevention and surveillance. The multiplier method is one of the most common methods used in estimating the size of key populations. When there are multiple data sources available, multiple size estimates that share a data source are calculated and then averaged to arrive at a single estimate, which is problematic as this ignores the correlation introduced by the shared source. A proposed Bayesian hierarchical framework circumvents the issue of correlation, yielding a single estimate using all data sources. We apply this framework to data collected in eSwatini to estimate the size of both female sex workers and men who have sex with men at the regional level. We also conduct simulations to assess how the hierarchical framework performs compared to the commonly used methods. In the simulation of the female sex worker analysis we find that 95 % confidence intervals for both methods perform similarly, with coverage proportions ranging from 91 % to 98 %. The hierarchical framework tended to be more precise as measured by the length of the confidence intervals. In the simulation of the men who have sex with men analysis we find that at one location, Mbabane, the hierarchical framework produces a confidence interval covering

the truth only 89 % of the time, versus 95% of the time using the average multiplier method. However, the hierarchical method produces a confidence interval with average length of 42 individuals for Nhlangano, as opposed to the average multiplier method which produces a confidence interval with average length of 168 individuals. This increase in precision may be linked to an increase in bias, as the average multiplier method MSM estimates seemed to have a smaller bias on average. We conclude that in the situation where there are multiple data sources available, the hierarchical framework may produce considerably more precise estimates, with the limitation that implementation of this method is potentially more challenging than calculating a simple average.

Thesis Committee

Primary Readers

Abhirup Datta (Primary Advisor)
Assistant Professor
Department of Biostatistics
Johns Hopkins Bloomberg School of Public Health

Stefan Baral
Associate Professor
Department of Epidemiology
Department of International Health
Department of Health Policy and Management
Johns Hopkins Bloomberg School of Public Health

Acknowledgments

Many thanks to Abhi Datta for giving me the opportunity to participate in this project, and for the help and guidance throughout. Special thanks are due to Stefan Baral for agreeing to be a reader and for providing valuable feedback.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

In the context of HIV prevention and surveillance, populations at high risk of acquiring and transmitting HIV are known as key populations. Understanding and reaching these key populations is a central part of strategies to end the HIV pandemic. As such, quantifying the size of these key populations in various locations is a problem that occupies governments, epidemiologists, and non-governmental organizations alike. However, this undertaking is complicated by the fact that individuals within these populations often face stigma and marginalization that prevent the use of standard methods for population size estimation in traditional demographic settings. There is a body of literature containing methods for estimating key population sizes, each method with its own strengths and weaknesses. Often in practice, multiple methods are used, which forces the researcher to grapple with some statistical difficulties. One such difficulty is how to unify these various estimates and their individual measures of uncertainty into a single estimate and uncertainty measure in a manner that is statistically sound. This thesis will first provide an overview of key population size estimation, followed by the application and evaluation

of a Bayesian hierarchical framework that solves some of the aforementioned statistical issues when using data obtained using the multiplier method.

1.1 Key Population Size Estimation

The global HIV pandemic itself needs no introduction. Efforts to combat the continued spread of HIV and to manage the care of infected individuals at the population level combine the efforts of governments, academic and public health institutions, and non-governmental organizations around the world. Two agencies of the United Nations, the World Health Organization (WHO) and the Joint United Nations Program on HIV/AIDS (UNAIDS) provide leadership and recommendations about how best to address the HIV epidemic. In 2016 the United Nations enacted a resolution to end AIDS by 2030 (Assembly, 2016), and with this resolution UNAIDS highlights three commitments: ensuring that 90% of people with HIV know their HIV status and utilize antiretroviral therapy, eliminating new HIV infections among children, and ensuring access of prevention options to 90% of people (UNAIDS, 2018). Specifically, UNAIDS emphasizes that prevention options need to be accessed by key populations (KP) including female sex workers (FSW) and their clients, men who have sex with men (MSM), transgender people, people who inject drugs, and prisoners. These populations often have the highest prevalence of HIV in countries with both concentrated and generalized HIV epidemics, and thus are a primary target for program interventions. Accurate KP size estimates allow organizations to efficiently allocate resources and are important for both monitoring the progress of interventions and modelling

the progression of the HIV epidemic (World Health Organization, 2010). However, due to the stigmatized or illicit nature of belonging to these populations, traditional surveillance systems for population size estimation are inadequate for KP. In recognition of this dilemma, the WHO has released guidelines and methods for reaching and estimating the size of these populations. Below we list these methods along with a brief description and citation before providing a more in depth overview of the multiplier method.

- Census/Enumeration: Census involves counting every individual within the key population. It requires that one has a complete list of all areas in which members of the key population congregate. Enumeration is similar to census, but treats the complete list of locations as a sampling frame. Sites are then randomly selected from the sampling frame, and then individuals are counted at the chosen sites. The final estimate is obtained by scaling the number counted according to the size and structure of the sampling frame. (Altaf et al., 2012) (Comiskey et al., 2013)

- Capture-Recapture: The most basic procedure of capture-recapture consists of going to a location and "tagging" all of the members of the key population by giving them a distinct item and counting everyone who received the item. Later, the researcher returns to this location and counts all of the members once again, being sure to note how many of them were "tagged" in the initial visit. The population estimate for that location is then calculated using the Lincoln-Petersen estimator using the the counts of both visits and the number of individuals who were counted twice. (Kimani et al., 2013)

- Multiplier Method: The multiplier method is a methodological cousin of the capture-recapture method. As an initial count of "tagged" individuals the multiplier method begins with the number of key population members who received services from a clinic (an STI or needle exchange clinic for example). Later, a survey of the key population is conducted, and within that survey individuals are asked if they received the service. The total estimate is then calculated again using the Lincoln-Petersen estimator. (Paz-Bailey et al., 2011,Johnston et al., 2011, Khalid et al., 2014)

- Network Scale Up: A survey of the general population is conducted in which respondents are asked how many individuals within a given key population they know. The respondents are also asked how many total people they know (total network). The estimate of the key population size is then obtained by taking the average proportion of known key population members to the total network size. (Ezoe et al., 2012, Salganik et al., 2011)

Each of these methods pose unique challenges and limitations. The census and enumeration methods have the advantage of a simple design and easily intelligible results, but can be costly to implement. In addition, they require that members of the key population congregate at known locations and are easily identifiable. In the case where a sizable fraction of the key population cannot be seen at any given location, census and enumeration methods will underestimate the size of the population. Similarly, the capture-recapture and multiplier methods require the assumption that each population member has an equal chance of selection, which is easily violated if some members of the

population of interest are hidden compared to others. The network scale up method has the strength that it can be conducted as part of a larger survey of the general population, but can suffer from bias if the key population is marginalized.

1.2 Multiplier Method

Of the methods mentioned above, the multiplier method is most commonly used to estimate key population sizes (Abdul-Quader, Baughman, and Hladik, 2014). The traditional capture-recapture method requires that two surveys of the key population are conducted (Petersen, 1896, Lincoln, 1930). In contrast, the multiplier method replaces one of the surveys with the count of key population members obtained from a clinic offering services or some similar source. Later, key population members participating in a survey are asked if they received these services. An estimate of the total key population size can be calculated from both the multiplier method and capture-recapture method in the same way, but the multiplier method has the advantage of requiring only a single survey. These surveys are often conducted using methods that offer incentives (Heckathorn, 1997, Johnston et al., 2013,Fearon et al., 2017) or make use of KP networks and KP meeting places (Rao et al., 2017, Weir et al., 2005).

In practice, it is very common that survey participants are asked about their attendance at multiple clinics or events which introduces an interesting statistical problem. If one knows both the marginal totals of each multiplier source and the number of individuals that participated in at least one source,

then a single key population size estimate can be obtained as described by (Darroch, 1958). However, with a single survey one cannot calculate the number of individuals who participated in at least one source. An approach taken in this situation is to calculate several estimates using the Lincoln-Petersen estimator, and then use the median or mean of these estimates (Holland et al., 2016). For the remainder of this thesis we will refer to this practice as the average multiplier method. From a statistical perspective, calculating either the mean or the median of several estimates ignores the correlation due to the shared survey source and is thus unsound. A model based solution to this proposed by Datta, 2019 incorporates a Bayesian hierarchical framework to produce a single estimate for each region. The Bayesian hierarchical framework and the average multiplier method are applied in the analysis section to compare the results they yield and their performance under simulation.

1.3 Context and Data: The HIV Epidemic in eSwatini

Among individuals aged 15 to 49, the prevalence of HIV in eSwatini is one of the highest in the world. Within this age group, over 1 in 4 individuals are living with HIV (CDC, 2013). Recent surveillance efforts show that within eSwatini progress toward UNAIDS 90-90-90 goals is being made. 90% of HIV positive individuals know their HIV status, and among these individuals about 85 % of them are currently using ART (UNAIDS, 2017).

As has been discussed, reaching key populations within eSwatini will be

crucial to continued progress against the HIV epidemic. While key populations can be difficult to reach in any nation, in eSwatini both sex work and sexual relations between members of the same sex are illegal and members of both FSW and MSM populations have reported discrimination, violence, and social exclusion that can prevent them from receiving the services they need (Kennedy et al., 2013, Logie et al., 2019, Risher et al., 2013). Thus, methods that can quickly estimate the size of these populations while maintaining their anonymity are of great importance.

To date, there has been only one size estimation effort for these populations in 2014 (Rao et al., 2017). It is this data that we use within the analysis portion of the thesis, and we discuss it's structure in the following section.

1.3.1 Data

The full details and methods regarding data collection can be found in (Rao et al., 2017). Using a modified version of the PLACE method and snowball sampling, two separate surveys were conducted, one for FSW and one for MSM. Respondents were required to be 18 years of age or older to participate.

In total, 532 MSM and 781 FSW were surveyed at 5 site locations and the surrounding area. In table 1.1 we summarize the counts and locations of the surveyed individuals, mapping the surrounding locations to one of the 5 sites. A more detailed table showing the exact locations of the respondents is available in section 2.3.

Participants in the survey were asked about their participation in events or receipt of objects or services, allowing for the use of the multiplier method.

	Location	Region	Count		Location	Region	Count
	Piggs Peak	Hhohho	57		Piggs Peak	Hhohho	127
	Mbabane	Hhohho	223		Mbabane	Hhohho	255
	Manzini	Manzini	177		Manzini	Manzini	257
MSM	Nhlangano	Shiselweni	70	FSW	Nhlangano	Shiselweni	47
	Lavumisa	Shiselweni	1		Lavumisa	Shiselweni	88
		Lubombo	4			Lubombo	7
		Total	532			Total	781

Table 1.1: Location distribution of the survey participants.

Below we describe each of these data sources, which will be referred to going forward as listings.

- Coupon: MSM were given coupons redeemable for HIV services in Mbabane and Manzini.
- Unique object identifier (UID): KP members at every site were given a deck of cards designed for this purpose.
- Rainbow Night: The rainbow night was a social event for MSM in Nhlangano.
- FLAS: FLAS is an acronym for the Family Life Association of eSwatini, an NGO that provides services for HIV prevention and sexual health. FSW in Manzini and Mbabane were asked if they attended a FLAS clinic in the month of September.

Broken down by location and KP, the structure of this data in the context of the multiplier method is shown in figures 1.2 and 1.3. We note two situations that either preclude the calculation of a single estimate or require that some data be discarded, exemplified by the coupon and FLAS data in Mbabane

and Manzini and the rainbow night MSM data collected in Nhlangano. These issues are labelled "incompleteness" and "misalignment" in Datta, 2019, and we discuss them in further detail in the following section.

1.3.2 Incompleteness and Misalignment

Incompleteness refers to the fact that the counts in some of the partitions of the Venn diagrams are unknown, specifically in the case of the MSM data in Nhlangano shown in figure 1.2. Note that the marginal totals of each circle are known (70, 12, and 106), but that there are three partitions for which we do not know the counts. From left to right, these partitions are the number of MSM in Nhlangano who attended the rainbow night but participated neither in the survey nor received the UID, the number of MSM who attended the rainbow night, participated in the survey and received the UID, and the number of MSM who received only the UID but did not participate in the survey and did not attend the rainbow night. In the multiplier method setting, because there is only one survey conducted, data will always be incomplete when there is more than 1 listing. With the average multiplier method, two Lincoln-Petersen estimates can be calculated for the MSM population in Nhlangano, one using the overlap between UID and the survey, and another using the overlap between rainbow night and the survey. Taking an average of the two estimates and confidence intervals ignores the fact that the estimates are correlated as they share a common data source, the survey. One advantage of the Bayesian hierarchical model based approach developed by Datta, 2019 is that it produces a single estimate for each region and thus circumvents the

issue of correlation.

Misalignment is exemplified by Manzini and Mbabane in the case of both MSM and FSW. Notice that for MSM we know that there were 105 total coupons distributed in Manzini and Mbabane, but we do not know how many were distributed within each city. For FSW we know that 186 total individuals attended FLAS clinics in September, and that at least 70 of these were in Mbabane, but we again do not the marginal totals in each city. Since these marginal totals are unknown, the Lincoln-Petersen estimate using the overlap between the coupons or FLAS and the survey cannot be calculated. This is a second advantage of the model based approach: rather than discarding the coupon and FLAS data, it can be integrated into the model and inform the final estimates.

1.3.3 Analysis Objectives

With this data in hand we have three objectives for the analysis. First, we want to use the Bayesian hierarchical framework to obtain a single KP size estimate and proportion for each region of eSwatini by generalizing the estimate obtained at each survey location to the region as a whole. A map of eSwatini, its regions, and the survey locations are shown in figure 1.1. Part of this objective involves estimating KP size in Lubombo despite their being very sparse data collected in that region.

The Bayesian hierarchical framework is heavily model based and includes choices of prior distributions and their associated parameters. As such, the second objective of the analysis is to determine how sensitive the KP size

Figure 1.1: Map of eSwatini showing its 4 regions. Data collection centers are labeled with text and circled with a green circle. Mbabane and Manzini are enclosed in a grey box to signify that some listings were conducted in both locations.

estimates are to these modeling choices. Finally, the third objective is to evaluate the performance of the Bayesian hierarchical framework and to compare its performance to that of the average multiplier method.

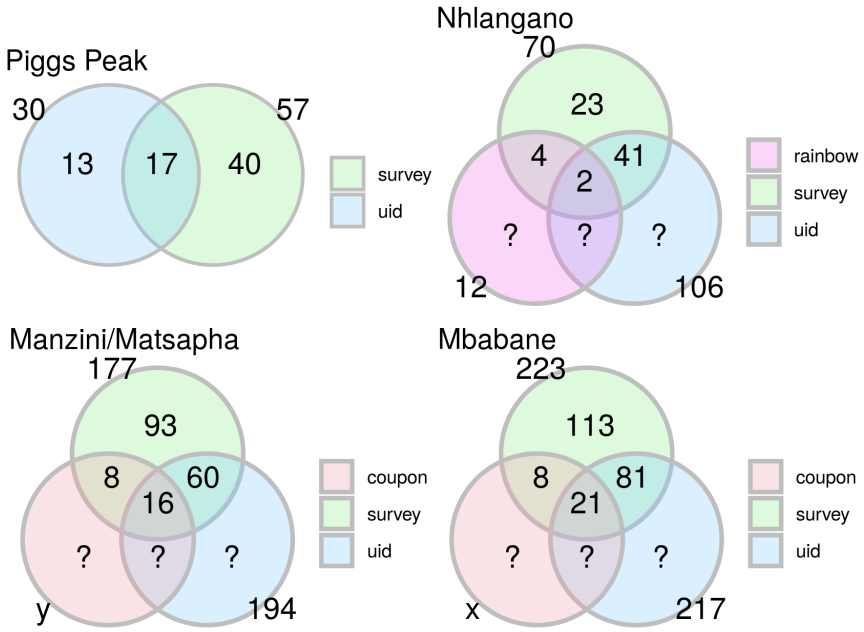

Figure 1.2: MSM counts and overlaps for the various data sources in each site. Note that the marginal total numbers of coupons distributed in Manzini and Mbabane (y and x) are not known, but the total number of coupons distributed is known, such that $y + x = 105$.

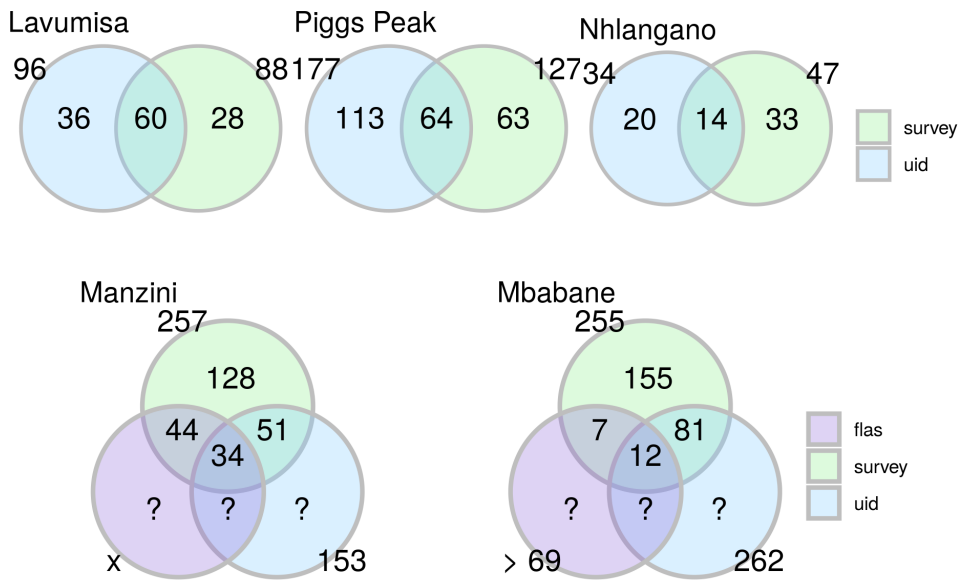

Figure 1.3: FSW counts and overlaps for the various data sources in each site. We know that the total number of FLAS attendees in Manzini and Mbabane is 186 and also that at least 70 of these were located in Mbabane.

Chapter 2

Methods

The remainder of the thesis is more narrow in focus than the introduction and will focus on an analysis of data collected within eSwatini. This section is divided into three sections. The first section describes the application of the Bayesian hierarchical framework developed by Datta, 2019 and explicitly defines both the Lincoln-Petersen estimator and the average multiplier method. The second section describes the simulation and sensitivity analyses conducted. The final section contains the details regarding the calculation of Tinkhundla population sizes in 2014, as well as the size of the populations that potentially could have participated in the survey.

2.1 Bayesian Hierarchical Framework

In section 1.3 context is provided for the multiplier method data collected within eSwatini. Recall that within each location of data collection, there are one or two KP counts in addition to a survey. The strength of the Bayesian hierarchical framework is that it integrates each of the locations and sources

such that a single estimate for each location can be obtained by using all of the data together. In obtaining an estimate for the KP proportion at the regional level we make the assumption that the Tinkhundla proportion is similar to the KP proportion within the survey location.

To begin, define N_i to be the KP size in the ith location, where $i = 1, 2, ...l$ is an index that denotes the location in which data has been collected and l is the total number of locations. We are primarily concerned with the estimation of the KP proportion in location i, or $\phi_i = \frac{N_i}{P_i}$, where P_i is the total population in that location in the same age range as the survey participants. It is this ϕ_i at a single location that we then generalize from a single location to the region as a whole. More specifically, if we let ϕ_R be the KP proportion at region R, then $\phi_R = \phi_i * P_R$, where P_R is the total population in the region. The top level of the hierarchical model then consists of N_i and its specified prior distribution. In this framework two priors are used: the Jeffrey's Prior $\frac{1}{N_i}$ and Binomial(P_i, ϕ_i) prior, both of which lead to conjugacy. When using the Binomial prior a Beta prior is also placed on ϕ_i of the form $\phi_i \sim Beta(\alpha_\phi, \beta_\phi)$, which makes the assumption that the KP proportion at each site is drawn from a common distribution.

The next level in this hierarchical model is concerned with the KP counts obtained from each listing and survey. Let n_{ij} be the count of KP members that participated in listing j at location i, where $j = 1, 2, ...s$ is an index that denotes the specific listing and s is the total number of listing or survey sources. In this framework each n_{ij} is modeled as a draw from Binomial(N_i, p_{ij}), where p_{ij} is the probability that a KP member in location i participates in the jth listing.

Each of these p_{ij} inclusion probabilities are modelled with a Beta(α_j, β_j) prior. We place a flat $\alpha_j = \beta_j = 1$ prior in the analysis, but evaluate how sensitive the results are to this choice in section 2.2.

The model and prior choices here all lead to conjugacy and can be sampled using a Gibbs sampler as described in Datta et al. 2019. We created 10 chains using the Gibbs Markov Chain Monte Carlo (MCMC) algorithm, each of length 10,000. The first 5,000 iterations from each chain were discarded as burn-in, and the remaining 5,000 iterations from each chain were combined as draws from the posterior distribution for each parameter. We checked the convergence of each chain using the Gelman Rubin diagnostic. The posterior mean of each N_i was then divided by P_i to obtain ϕ_i. Information about how P_i was calculated is contained in section 2.3. To obtain ϕ_R For Lubombo, the region for which we had no usable data, we used the posterior mean for $\phi_i = \alpha_\phi / (\alpha_\phi + \beta_\phi)$.

For both MSM and FSW we have data from two locations in the Hhohho region, Pigg's Peak and Mbabane. For FSW we have data from Nhlangano and Lavumisa in the Shiselweni region. Thus, in Hhohho we first multiplied ϕ_i from Pigg's Peak by the total population in the northern Tinkhundlas (Pigg's Peak, Ndzingeni, Mandlangempisi, Ntfonjeni, Timphisini, Mayiwane, Mhlangatane). Then, we multiplied the ϕ_i for Mbabane by the total population in the southern Tinkhundlas (Nkhaba, Maphalaleni, Motjane, Mbabane East, Mbabane west, Hhukwini, Lobamba). This yields a northern and southern MSM or FSW size estimate, which we summed to obtain a single estimate for all of Hhohho. We used the same process to obtain a single estimate for

Shiselweni, except we divided by east (Matsanjeni South, Somntongo, Hosea) and west (Sigwe, Ngudzeni, Zombodze, Emuva, Maseyisini, Mbangweni, Mtsambama, Kubuta, Nkwene, Gege).

2.1.1 Average Multiplier Method

For comparison, we also estimate N_i at each location using the average multiplier method. Let n_1 be the number of individuals that participated in the survey at location i and let n_2 be the number of individuals that participated in another listing at location i. Survey respondents are asked if they participated in this other listing, and we thus have m the number of individuals who participated in both the survey and the listing. The Lincoln-Petersen estimate $\hat{N}_i$ is defined as

$$\hat{N}_i = \frac{n_1 n_2}{m}$$

with variance

$$Var(\hat{N}_i) = \frac{n_1 n_2 (n_1 - m)(n_2 - m)}{m^3}.$$

We estimate a 95% confidence interval for $\hat{N}_i$ using a normal approximation given by

$$\hat{N}_i \pm 1.96\sqrt{Var(\hat{N}_i)}.$$

The average multiplier method is used when there are multiple listings in a location in addition to the survey. In this case, a separate Lincoln-Petersen

estimate is calculated using the overlap between each listing and the survey, and then these estimates and confidence intervals are averaged. We used this method to calculate an estimate $\hat{N}_i$ and confidence interval for the MSM data in Nhlangano, for comparison to the Bayesian hierarchical framework.

2.2 Simulation and Sensitivity Analyses

The goal of the simulation analysis is to verify two things: to examine how well the Bayesian framework and MCMC method performs and to compare its performance to the average multiplier method. The sensitivity analysis aims to evaluate how sensitive the results are to changes in prior specification.

2.2.1 Simulation

Using the posterior mean estimates obtained in the analysis described in section 2.1, we performed the following procedure 100 times. We first simulated a new N_i for each location by taking a random draw from a Binomial(P_i, ϕ_i) distribution. We note that in doing so we are in effect violating the assumption that each ϕ_i is drawn from a common distribution, and thus this simulation also serves to evaluate the sensitivity of the analysis to this assumption. With the N_is in hand, we then simulate a new data set of the observed data by drawing a count for each of the data sources from a Multinomial(N_i, p_1, p_2,...,p_s) distribution. We then proceed from this new simulated data and perform the analysis described in section 2.1 with the addition of calculating estimates and confidence intervals using the average multiplier method. In the case of the MSM data in Nhlangano, we calculate one estimate for the UID source

Table 2.1: MSM : Region-specific age-group proportion, annual growth rates, counts and weights used for reported locations assigned to each Tinkhundla in determining the total population representing the survey at each site.

Region	Percent of Males (18- 32)	Annual Growth Rate	Site	Locations reported	Counts	Tinkhundla	Weight
Manzini	30 %	1.1 %	Manzini	Manzini/Matsapha	173	Manzini North and South, Kwaluseni	173/177
				Malkerns	3	Lobamba Lombdzala	4/177
				Mahlanya	1		
Hhohho	29 %	1.3 %	Mbabane	Mbabane/Ezulwini	155	Mbabane East and West, Lobamba	155/223
				Ngwenya	61	Motjane	68/223
				Oshoek	3		
				Motshane	4		
			Piggs Peak	Piggs Peak	54	Piggs Peak	54/ 57
				Matsamo	3	Timphisini, Ntfonjeni	3/57
Shiselweni	26 %	-0.2 %	Nhlangano	Nhlangano	70	Mbangweni	1

and one for the coupon source, and then average both the point estimate and the confidence intervals to arrive at the final estimate. At each iteration, we check if the average multiplier 95% confidence interval and model based 95% credible interval contains the true N_i.

2.2.2 Sensitivity Analysis

Recall that in section 2.1 we used $\alpha_i = \beta_i = 1$ as parameters in the beta prior placed on each inclusion probability p_i. To assess how sensitive the results are to this specification, we performed the same analysis using three alternate values, $0, 0.5$, and 1.5.

2.3 2014 Population Size Calculations

In this section we provide detail regarding the calculation of the total number of individuals in the same age range as those who participated in the survey, defined as P_i. As a source for calculating P_i we made use of the preliminary results of the 2017 eSwatini census, which contained population size data

Table 2.2: FSW : Region-specific age-group proportion, annual growth rates, counts and weights used for reported locations assigned to each Tinkhundla in determining the total population representing the survey at each site.

Region	Percent of Females (18- 35)	Annual Growth Rate	Site	Locations reported	Counts	Tinkhundla	Weight
Manzini	35 %	1.1 %	Manzini	Manzini/Matsapha	249	Manzini North and South, Kwaluseni	249/257
				Malkerns	8	Lobamba Lombdzala	8/257
Hhohho	32 %	1.3 %	Mbabane	Mbabane/Ezulwini	198	Mbabane East and West, Lobamba	198/255
				Ngwenya	55	Motjane	57/255
				Oshoek	2		
			Piggs Peak	Piggs Peak	121	Piggs Peak	121/127
				Matsamo	2	Timphisini	2/127
				Buhleni	4	Mayiwane	4/127
Shiselweni	29 %	-0.2 %	Nhlangano	Nhlangano	47	Mbangweni	1
			Lavumisa	Lavumisa	81	Somntongo	81/88
				Hluthi	3	Hosea	3/88
				Matsanjeni	4	Matsanejni South	4/88

only at the regional and Tinkhundla level (Central Statistical Office, 2017). Recall that the data used in this analysis was collected in eSwatini in 2014 using venue based snowball sampling. Thus, there are two issues to address in using the census data. The first issue is that that while we have grouped respondents to one of five locations, an individual surveyed in the city of Mbabane for example may not live in Mbabane proper but rather in one of the surrounding Tinkhundlas. Since the survey included information about where the respondent lives, we used a weighted average of the Tinkhundla populations based on the proportion of respondents who reported living in each Tinkhundla. For example, if we had a total of 100 individuals surveyed in Mbabane, with 4 reporting that they reside in the Lobamba Tinkhundla, then we multiplied the population of Lobamba by 4/100, repeated this for each of the reported Tinkhundlas, and then summed these numbers. In a few cases we also included a weighting of Tinkhundla populations for which no respondents reported being residents due to that Tinkhundla's geographical proximity to the survey site. The weights and Tinkhundlas included in this

calculation are visible in tables 2.1 and 2.2.

The second issue is that the multiplier method data was collected in eSwatini in 2014, whereas the census data was collected in 2017. To address this we performed a simple interpolation by calculating an average yearly population change rate from 2007 to 2017. The data from 2007 was only available at the regional level, and thus we make the assumption that the rate of change at the Tinkhundla level was similar to that of the region as a whole. We applied this interpolation to the weighted population numbers described in the previous paragraph.

Finally, because P_i is a total population of individuals of similar age to those who were surveyed we calculated the 90th percentile of age in the survey respondents, separately for MSM and FSW. Survey respondents were required to be 18 years of age to participate, so we took 18 to the 90th percentile as our target age range. Again using the 2017 census data, we calculated the proportion of the population within this age range for each region. We arrived at the final population numbers by multiplying the interpolated and weighted population sums described in the previous paragraphs by this regional proportion.

Chapter 3

Results

We divide the results chapter into sections. The first section contains regional population size estimates for both the MSM and FSW analyses, using both the framework described in Datta, 2019 and the average multiplier method. The second section shows the results of the same analyses conducted using different prior choices and thus demonstrates the robustness of the results to prior specification. The third section contains the results of the simulations that show the performance of both the hierarchical framework and the average multiplier method.

3.1 Key Population Size Estimates

The regional key population size estimates for FSW and MSM are contained in tables 3.1 and 3.2. There is no estimate available for the Lubombo region using the average multiplier method as there was insufficient data. For the other regions the two methods provide similar estimates.

Table 3.1: Region specific size estimates of FSW in the age group of $15-49$ years. AMM refers to the average multiplier method and model refers to the Bayesian hierarchical model approach. The values in parentheses are 95% confidence intervals.

Region	AMM FSW size estimate	Model FSW ize estimate
Hhohho	6393 (5528, 7259)	6415 (5464 - 7493)
Manzini	2284 (2020, 2549)	2114 (1872 - 2377)
Shiselweni	2007 (1605, 2408)	2052 (1568 - 2722)
Lubombo		4059 (2045 - 8328)

Table 3.2: Region specific size estimates of MSM in the age group of $15-49$. AMM refers to the average multiplier method and model refers to the Bayesian hierarchical model approach. The values in parentheses are 95% confidence intervals.

Region	AMM MSM Estimate	Model MSM estimate
Hhohho	3509 (2940, 4080)	3465 (2888 - 4176)
Manzini	2649 (2298, 3000)	2650 (2281 - 3068)
Shiselweni	1750 (1188, 2312)	1869 (1544 - 2238)
Lubombo		2015 (1451 -2764)

Region	Male Population 15-49	Female Population 15-49
Hhohho	85918	84784
Manzini	95720	100360
Shiselweni	45216	48391
Lubombo	52519	52762

Table 3.3: Regional populations aged 15-49

3.2 Sensitivity to Prior Specification

Prior specification refers to two situations. The first is that two alternative priors were placed on N_i, a Jeffrey's prior,$N_i \sim \frac{1}{N_i}$, and a Binomial prior $N_i \sim$ Binomial(P_i, ϕ_i). In figure 3.1 one can see that the two choices provide very similar estimates, though it seems that there is more variation in the results obtained using the Jeffrey's prior. The estimates reported in the section above were obtained using the Binomial prior, as it permits the posterior mean for ϕ_i to be used as estimates for the KP proportions in Lubombo.

The second situation is that the same Beta prior was placed on the inclusion probability for each listing. The results reported in the previous section were obtained using $\alpha = \beta = 1$, but here we evaluate the sensitivity of the results to changes in the specification of α and β. Figure 3.2 demonstrates how the estimates change when the values are changed from 1 to 0, 0.5, and 1.5. Again we see that the estimates do not appreciably change.

3.3 Simulation Analysis

3.3.1 Female Sex Workers

In Figures 3.4 and 3.5 we compare the performance of the two methods under simulation, first by looking at the coverage performance of the confidence intervals obtained using both methods, and then comparing the precision of the estimates and intervals. For FSW, the performance of the two methods is very similar, with coverage probability ranging from 91 % to 98 %. In regard to precision we show the average length of confidence intervals, average

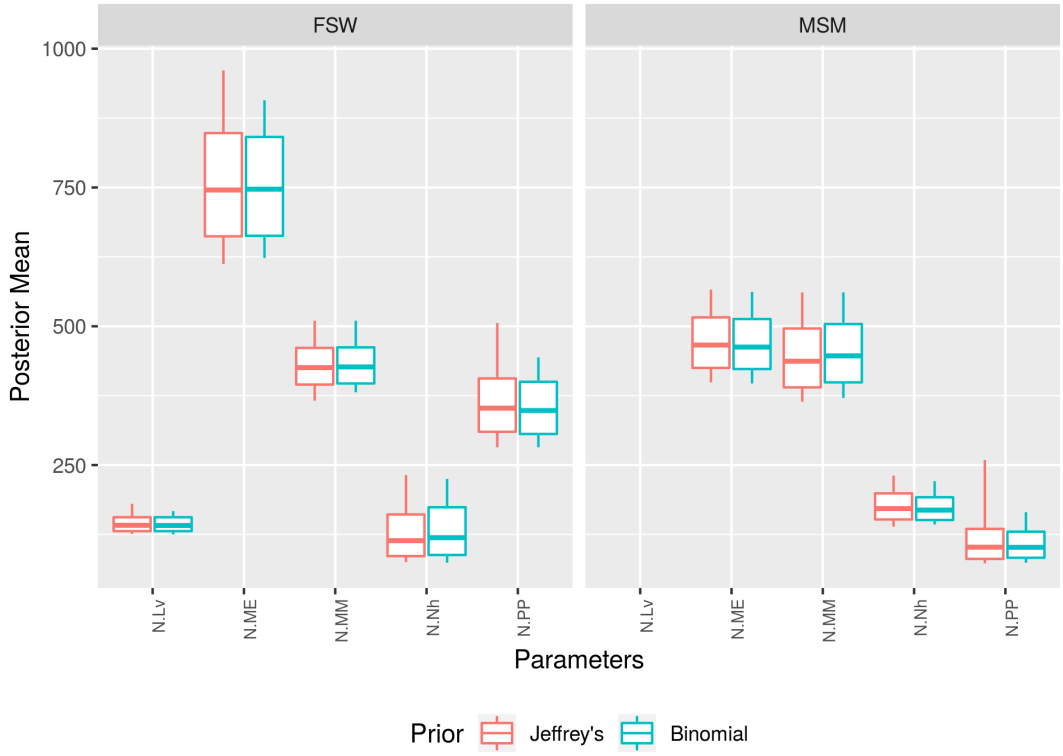

Figure 3.1: A side by side comparison of KP size estimates at each location with one of two priors placed on N_i, the KP size

absolute distance from the truth, and average bias. Again both methods perform similarly, with the model based approach providing more precision and less bias in Mbabane and Manzini, but the average multiplier method providing more precision and less bias in Nhlangano.

3.3.2 Men Who have Sex with Men

For MSM, the comparison of the performance of the methods is shown in tables 3.6 and 3.7. Performance of the confidence intervals obtained from the two methods is again similar, although the interval for Mbabane using the Bayesian hierarchical model performs noticeably worse (89 % coverage

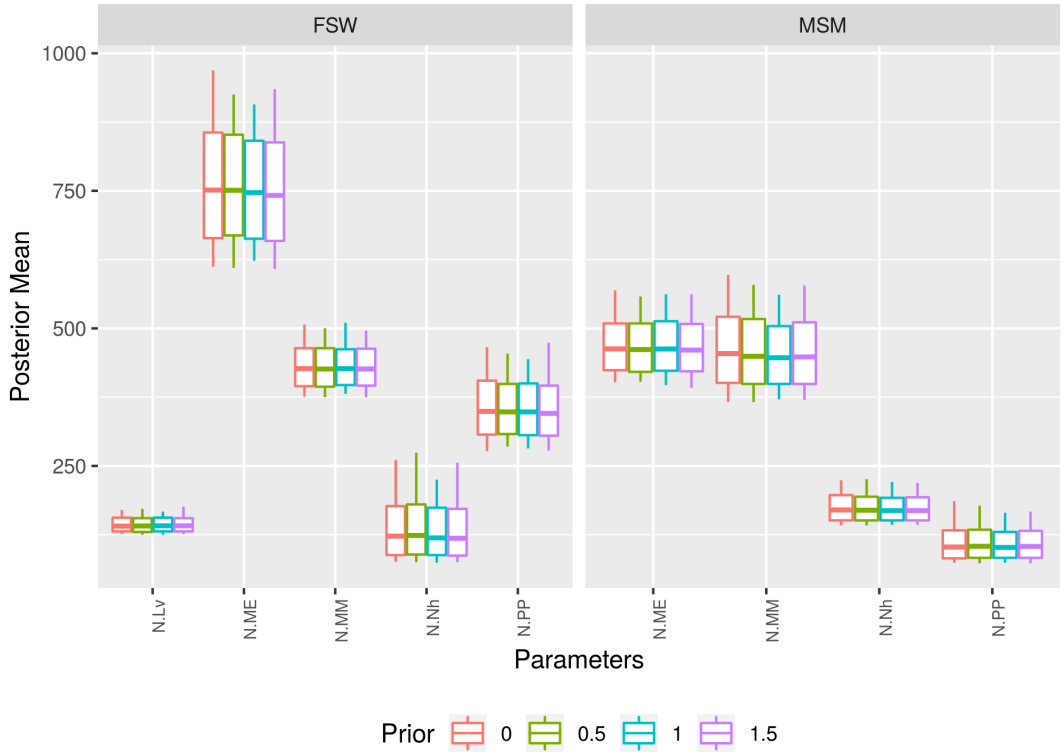

Figure 3.2: Side by side comparisons of the KP size estimates using 4 values for $\alpha_i = \beta_i$ in the Beta prior placed on the inclusion probabilities for participating in each listing.

probability) and compared to the average multiplier estimate (95 % coverage probability). In terms of precision however, the model based approached seems to perform markedly better in the case of Nhlangano, where the average length of a confidence interval is 42 units using the Bayesian framework as opposed to 168 units using the average multiplier method. However, the average multiplier method seems to provide noticeably less bias, with a smaller average bias at every location except for Nhlangano.

Table 3.4: FSW size estimation simulation results for both the Bayesian hierarchical model approach and the average multiplier method. True values and coverage probabilities were averaged over 100 simulations. The estimates for the model approach are the posterior mean. The true values are averaged over the 100 simulations, as each simulation uses a different true value. CP denotes coverage probability.

	Region	True Pop Size	Pop Size Estimate	CP	Average Length of Confidence Interval
Model Based FSW (18-32 years) size estimates	Piggs Peak	345	346	98 %	93
	Nhlangano	121	126	96 %	93
	Lavumisa	140	141	92 %	26
	Mbabane	751	751	94 %	189
	Manzini	430	432	93 %	80
Average Multiplier FSW (18-32 years) size estimates	Piggs Peak	345	349	98 %	97
	Nhlangano	121	121	91 %	83
	Lavumisa	140	141	93 %	26
	Mbabane	751	753	94 %	204
	Manzini	430	433	94 %	91

Table 3.5: A comparison of the precision of both the estimates and confidence intervals of the two methods using the FSW size estimation simulation results.

	Region	Average Distance From Truth	Average Bias
Model Based Approach	Piggs Peak	18	2
	Nhlangano	18	5
	Lavumisa	5	1
	Mbabane	37	-0.08
	Manzini	18	1
Average Multiplier Method	Piggs Peak	19	4
	Nhlangano	17	-0.54
	Lavumisa	5	0.88
	Mbabane	42	2
	Manzini	20	2

Table 3.6: MSM size estimation simulation results for both the Bayesian hierarchical model approach and the average multiplier method. True values and coverage probabilities were averaged over 100 simulations. The estimates for the model approach are the posterior mean. The true values are averaged over the 100 simulations, as each simulation uses a different true value. CP denotes coverage probability.

	Region	True Pop Size	Pop Size Estimate	CP	Average Length of Confidence Interval
Model Based MSM (18-32 years) size estimates	Piggs Peak	104	109	96 %	51
	Nhlangano	167	165	92 %	42
	Mbabane	459	449	89 %	84
	Manzini	451	470	94 %	129
Average Multiplier MSM (18-32 years) size estimates	Piggs Peak	104	109	93 %	61
	Nhlangano	167	173	96 %	168
	Mbabane	459	455	95 %	91
	Manzini	451	452	91 %	118

Table 3.7: A comparison of the precision of both the estimates and confidence intervals of the two methods using the MSM size estimation simulation results. Average distance is the average absolute distance from the truth.

	Region	Average Distance From Truth	Average Bias
Model Based Approach	Piggs Peak	10	5
	Nhlangano	9	-2
	Mbabane	20	-10
	Manzini	32	19
Average Multiplier Method	Piggs Peak	13	5
	Nhlangano	29	6
	Mbabane	19	-3
	Manzini	23	1

Chapter 4

Discussion

Reiterating one of the goals for the analyses in this thesis, we set out to estimate the size and proportion of key populations in eSwatini using a Bayesian hierarchical framework developed by Datta, 2019. This we achieved, and the estimates obtained using the Bayesian hierarchical model show close agreement to those obtained using the average multiplier method. In addition, the model approach allowed us to estimate the KP proportion in Lubombo, a region in Zambia for which data were too sparse to use the average multiplier method. There are alternatives methods to infer KP proportions in regions for which there are no data using random effect models and demographic data (Datta et al., 2017). However, this method suffers from a similar limitation as the average multiplier method, in that multiple estimates within a region are considered independent when in fact they usually share sources of data.

We also set out to evaluate the performance of both the model based method and the average multiplier method under simulation. Ostensibly, this simulation analysis would answer the question of whether or not one method performs better than the other. In this case, however, it is difficult to say. For

FSW, the coverage probabilities of the confidence intervals for both methods were very similar and the precision of the model based estimates seemed to be slightly better. The situation for MSM is more interesting. Overall, the performance of confidence intervals in terms of coverage probability was similar between methods, with the exception of Mbabane. We found that while the average multiplier confidence interval for Mbabane covered the truth 95 % of the time, the model based interval covered the truth only 89 % of the time. In terms of precision, the two methods also performed similarly, with the exception of Nhlangano, where the model based confidence interval is a fourth of the size of the average multiplier method interval. This is notable because Nhlangano was the only location that exemplified the incomplete situation for which the model based approach is a solution. In that sense, it does appear that the Bayesian hierarchical approach outperforms the average multiplier method when there are incomplete data. Nevertheless, this increase in precision may be linked to an increase in bias as compared to the average multiplier method, although the bias is fairly small when compared to the size estimates.

Stepping back from purely statistical considerations, it is also important to consider the context in which these estimates are being obtained. While it seems that the Bayesian approach offers some advantages, it also requires a more extensive statistical background to understand and calculate. This begs the question of whether the improvement of the model approach outweighs the cost of finding someone with the necessary skills to carry out the analysis. While a definitive answer to this question is outside the scope of this thesis,

we suggest that it depends on the specific data set in hand. It seems plausible that for a given location, the more cases in which there are incomplete and misaligned data the greater the precision of the model based approach relative to the average multiplier method. Currently, we are working on the development of an R package that can make conducting this type of analysis more accessible.

References

Assembly, UN (2016). "Political Declaration on HIV and AIDS: on the fast-track to accelerate the fight against HIV and to end the AIDS epidemic by 2030". In: *New York: United Nations*.

UNAIDS (2018). "Global Aids Monitoring 2019: Indicators for monitoring the 2016 Political Declaration on Ending AIDS". In:

World Health Organization (2010). *Guidelines on estimating the size of populations most at risk to HIV*. Geneva: World Health Organization. ISBN: 978-92-4-159958-0.

Altaf, Arshad, Ajmal Agha, MH McKinzie, Qamar Abbas, Salma Batool Jafri, and Faran Emmanuel (2012). "Size estimation, HIV prevalence and risk behaviours of female sex workers in Pakistan". In: *JPMA-Journal of the Pakistan Medical Association* 62.6, p. 551.

Comiskey, Catherine, Orla Dempsey, Danijela Simic, and Sladjana Baroš (2013). "Injecting drug users, sex workers and men who have sex with men: a national cross-sectional study to develop a framework and prevalence estimates for national HIV/AIDS programmes in the Republic of Serbia". In: *BMJ open* 3.5, e002203.

Kimani, Joshua, Lyle R McKinnon, Charles Wachihi, Judith Kusimba, Gloria Gakii, Sarah Birir, Mercy Muthui, Anthony Kariri, Festus K Muriuki, Nicholas Muraguri, et al. (2013). "Enumeration of sex workers in the central business district of Nairobi, Kenya". In: *PloS one* 8.1, e54354.

Paz-Bailey, G, JO Jacobson, ME Guardado, FM Hernandez, AI Nieto, M Estrada, and J Creswell (2011). "How many men who have sex with men and female sex workers live in El Salvador? Using respondent-driven sampling and capture–recapture to estimate population sizes". In: *Sexually transmitted infections* 87.4, pp. 279–282.

Johnston, Lisa, Ahmed Saumtally, Sewraz Corceal, Indrasen Mahadoo, and Farida Oodally (2011). "High HIV and hepatitis C prevalence amongst injecting drug users in Mauritius: findings from a population size estimation

and respondent driven sampling survey". In: *International Journal of Drug Policy* 22.4, pp. 252–258.

Khalid, Farhat J, Fatma M Hamad, Asha A Othman, Ahmed M Khatib, Sophia Mohamed, Ameir Kh Ali, and Mohammed JU Dahoma (2014). "Estimating the number of people who inject drugs, female sex workers, and men who have sex with men, Unguja Island, Zanzibar: results and synthesis of multiple methods". In: *AIDS and Behavior* 18.1, pp. 25–31.

Ezoe, Satoshi, Takeo Morooka, Tatsuya Noda, Miriam Lewis Sabin, and Soichi Koike (2012). "Population size estimation of men who have sex with men through the network scale-up method in Japan". In: *PloS one* 7.1, e31184.

Salganik, Matthew J, Dimitri Fazito, Neilane Bertoni, Alexandre H Abdo, Maeve B Mello, and Francisco I Bastos (2011). "Assessing network scale-up estimates for groups most at risk of HIV/AIDS: evidence from a multiple-method study of heavy drug users in Curitiba, Brazil". In: *American journal of epidemiology* 174.10, pp. 1190–1196.

Abdul-Quader, Abu S., Andrew L. Baughman, and Wolfgang Hladik (2014). "Estimating the size of key populations: current status and future possibilities". In: *Current Opinion in HIV and AIDS* 9.2, pp. 107–114. ISSN: 1746-630X. DOI: 10.1097/COH.0000000000000041. URL: http://content.wkhealth.com/linkback/openurl?sid=WKPTLP:landingpage&an=01222929-201403000-00004 (visited on 02/25/2019).

Petersen, Carl Georg Johannes (1896). "The yearly immigration of young plaice in the Limfjord from the German sea". In: *Rept. Danish Biol. Sta.* 6, pp. 1–48.

Lincoln, Frederick Charles et al. (1930). "Calculating waterfowl abundance on the basis of banding returns". In: *US Dept. of Agriculture*.

Heckathorn, Douglas D. (1997). "Respondent-Driven Sampling: A New Approach to the Study of Hidden Populations". In: *Social Problems* 44.2, pp. 174–199. ISSN: 00377791, 15338533. DOI: 10.2307/3096941. URL: https://academic.oup.com/socpro/article-lookup/doi/10.2307/3096941 (visited on 04/03/2019).

Johnston, Lisa G, Dimitri Prybylski, H Fisher Raymond, Ali Mirzazadeh, Chomnad Manopaiboon, and Willi McFarland (2013). "Incorporating the service multiplier method in respondent-driven sampling surveys to estimate the size of hidden and hard-to-reach populations: case studies from around the world". In: *Sexually transmitted diseases* 40.4, pp. 304–310.

Fearon, Elizabeth, Sungai T Chabata, Jennifer A Thompson, Frances M Cowan, and James R Hargreaves (2017). "Sample Size Calculations for Population Size Estimation Studies Using Multiplier Methods With Respondent-Driven Sampling Surveys". In: *JMIR public health and surveillance* 3.3.

Rao, Amrita, Shauna Stahlman, James Hargreaves, Sharon Weir, Jessie Edwards, Brian Rice, Duncan Kochelani, Mpumelelo Mavimbela, and Stefan Baral (2017). "Sampling key populations for HIV surveillance: results from eight cross-sectional studies using respondent-driven sampling and venue-based snowball sampling". In: *JMIR public health and surveillance* 3.4.

Weir, S, J Tate, SB Hileman, M Khan, and E Jackson (2005). "PLACE. Priorities for Local AIDS Control Efforts: a manual for implementing the PLACE method." In: *Measure Evaluation*.

Darroch, John N (1958). "The multiple-recapture census: I. Estimation of a closed population". In: *Biometrika* 45.3/4, pp. 343–359.

Holland, Claire E., Seni Kouanda, Marcel LouguÃľ, Vincent Palokinam Pitche, Sheree Schwartz, Simplice Anato, Henri Gautier Ouedraogo, Jules Tchalla, Clarence S. Yah, Laurent Kapesa, Sosthenes Ketende, Chris Beyrer, and Stefan Baral (2016). "Using Population-Size Estimation and Cross-sectional Survey Methods to Evaluate HIV Service Coverage Among Key Populations in Burkina Faso and Togo". In: *Public Health Reports* 131.6, pp. 773–782. ISSN: 0033-3549, 1468-2877. DOI: 10.1177/0033354916677237. URL: http://journals.sagepub.com/doi/10.1177/0033354916677237 (visited on 03/25/2019).

Datta, Abhirup (2019). "Bayesian multi-region population size estimation using incomplete and misaligned capture-recapture data". In: p. 40.

CDC (2013). "CDC in Swaziland". In: *CDC Factsheet*.

UNAIDS (2017). *Eswatini*. URL: http://www.unaids.org/en/regionscountries/countries/swaziland.

Kennedy, Caitlin E, Stefan D Baral, Rebecca Fielding-Miller, Darrin Adams, Phumlile Dludlu, Bheki Sithole, Virginia A Fonner, Zandile Mnisi, and Deanna Kerrigan (2013). "âĂIJThey are human beings, they are SwaziâĂİ: intersecting stigmas and the positive health, dignity and prevention needs of HIV-positive men who have sex with men in Swaziland". In: *Journal of the International AIDS Society* 16, p. 18749.

Logie, Carmen H, Lisa V Dias, Jesse Jenkinson, Peter A Newman, Rachel K MacKenzie, Tampose Mothopeng, Veli Madau, Amelia Ranotsi, Winnie Nhlengethwa, and Stefan D Baral (2019). "Exploring the Potential of Participatory Theatre to Reduce Stigma and Promote Health Equity for Lesbian,

Gay, Bisexual, and Transgender (LGBT) People in Swaziland and Lesotho". In: *Health Education & Behavior* 46.1, pp. 146–156.

Risher, Kathryn, Darrin Adams, Bhekie Sithole, Sosthenes Ketende, Caitlin Kennedy, Zandile Mnisi, Xolile Mabusa, and Stefan D Baral (2013). "Sexual stigma and discrimination as barriers to seeking appropriate healthcare among men who have sex with men in Swaziland". In: *Journal of the International AIDS Society* 16, p. 18715.

Central Statistical Office, eSwatini (2017). "The 2017 Population and Housing Census: Preliminary Results". In:

Datta, Abhirup, Wenyi Lin, Amrita Rao, Daouda Diouf, Abo Kouame, Jessie K. Edwards, Le Bao, Thomas A. Louis, and Stefan Baral (2017). "Bayesian estimation of MSM population in CÃť te d'Ivoire". In: *bioRxiv*. DOI: 10.1101/213926. URL: http://biorxiv.org/lookup/doi/10.1101/213926 (visited on 02/25/2019).

Andrew Pita

248 S Wolfe Street
Baltimore, MD 21231

Email: apita3@jhmi.edu
Phone: (714) 768-1425
https://github.com/andrewpita

Education

Johns Hopkins Bloomberg School of Public Health — May 2019
ScM Biostatistics

Loyola Marymount University, Los Angeles CA — 2014
B.S Biology, minor in Classics
Cum Laude

Experience

Graduate Research Assistant, Department of Biostatistics, Johns Hopkins
07/01/2018 – Present,
Advisor: Abhirup Datta

- Performed statistical analyses, modeling, and programming for an epidemiologic study that aims to estimate the size of populations at risk of HIV infection in countries with a high HIV burden
- Conducted simulations to assess the performance of the model and its sensitivity to assumptions
- Created publication quality tables, figures and performed quality assurance as these estimates will be used to allocate resources for HIV prevention
- Employed high quality programming practices and documentation to ensure reproducibility

Jr. Specialist, Brain Imaging Center,
08/2015-08/2017
Department of Psychiatry, University of California, Irvine
Supervisor: Dr. Joe Wu

- Created pipelines using Unix scripting, Python, and Matlab to preprocess and analyze PET and DTI brain images from subjects with traumatic brain injury and other neurological disorders
- Utilized the Sun Grid Engine system to schedule and run batch jobs on a high performance computing cluster
- Designed experiments to compare subjects to controls and to evaluate differences between different software packages and versions
- Trained and supervised 4 undergraduates and new additions to the lab and ensured that the work done by all lab members was appropriately documented

Chief Medical Scribe, Scribe America at Congress Orthopedic Associates, Pasadena CA
02/2015-08/2015
Supervisor: Emmy Ogawa

- Interfaced between the implementation team at MD Suite (EMR company) and the group doctors to tailor the electronic medical record (EMR) system to fit the needs of the doctors and scribes
- Managed three other scribes on site by creating the coverage schedule, reviewing their charts, and communicating with the physician to ensure the quality of the site
- Communicated with Scribe America upper management and created monthly progress reports for the site

Research Assistant, Biology Department — 01/2014 – 06/2014
Loyola Marymount University, Los Angeles CA
Supervisor: Dr. Kam Dahlquist

- Created and ensured the quality of genomic databases
- Performed statistical analyses on microarray data and performed pathway analyses of expression levels

- Documented protocol and necessary files on an online wiki
- Communicated project status weekly to other research assistants and head professors to create a plan for future work to ensure that the team worked cohesively to accomplish its goals

Teaching Assistant, Biology Department 08/2013-06/2014
General Biology Lab
Loyola Marymount University, Los Angeles CA

- Provided four hours of instruction weekly to twelve students regarding lab techniques, the writing and presenting of scientific information, and the use of excel and statistical analysis tools
- Met with the other teaching assistants and the head professors bi- weekly to deliver the best possible educational experience

Programming Languages

- Python
- R, Shiny
- Linux/Unix scripting
- HPC cluster
- SQL

Research Publications and Conference Presentations

- **Size Estimation of Key Populations in the HIV Epidemic in eSwatini Using Incomplete and Misaligned Capture-Recapture Data,** Abhirup Datta, **Andrew Pita,** Amrita Rao, Bhekie Sithole, Zandile Mnisi, Stefan Baral, 2019, submitted to the Annals of Applied Statsitics
- **Altered Fractional Anisotropy in Cerebellum of Individuals with Fetal Alcohol Spectrum Disorders vs. Controls vs. Traumatic Brain Injury,** Fareshte Erani, **Andrew Pita**, Joseph Wu, 2017 *Society of Behavioral Medicine*
- **High Sensitivity and Specificity in Brain Injury Diagnostic Method Using Statistical Parametric Mapping of Resting Positron Emission Tomography, Andrew Pita**, Eric Chang, Cassidy Nguyen, Joseph Wu, *2016 North American Brain Injury Society*
- **Access to ants diversifies insect visitation to Castor bean (Ricinus communis L.) inflorescences in southern California,** Colin Ehret, Julie Hernandez , **Andrew Pita** , and Victor Carmona-Galindo, *BIOS.*
- **Whole genome comparison of a large collection of mycobacteriophages reveals a continuum of phage genetic diversity**, Contributor as part of the HHMI SEA Phage Program, Pope WH, Bowman CA, Russell DA, et al. Kolter R, ed. *eLife*. 2015.
- **Discovering Mycobacteriophage that Infects *Mycobacterium smegmatis*, Andrew Pita**, *2011 Loyola Marymount Undergraduate Research Symposium*
- **Constructing a GenMAPP-compatible gene database for *Streptococcus pneumoniae* to perform a pathway analysis on microarray data comparing biofilm versus planktonic forms**, **Andrew Pita**, *2014 Loyola Marymount Undergraduate Research Symposium, and 2014 Beta Beta Beta Biological Honor Society's Pacific District Convention*

Professional Affiliations

Sigma Xi Research Society
Eta Sigma Phi Classics Honor Society

References

Abhirup Datta, PhD
Assistant Professor
Department of Biostatistics

Johns Hopkins Bloomberg School of Public Health
615 N Wolfe Street, Room E3640
Baltimore MD 21205
(410) 502-2988
Email: abhidatta@jhu.edu

Joseph Wu, MD
Professor Emeritus, Former Clinical Director for Brain Imaging Center
University of California, Irvine, School of Medicine
199 Irvine Hall
Irvine, CA 92697
(949) 824-7867
Email: jcwu@uci.edu

Kam Dahlquist, PhD
Professor of Biology
Department of Biology
Loyola Marymount University
1 Loyola Marymount University Dr
Los Angeles, CA 90045
(310) 338-7697
Email: Kam.Dahlquist@lmu.edu

Matt Dillon, PhD
Professor of Classics
Department of Classics and Archaeology
Loyola Marymount University
1 Loyola Marymount University Dr
Los Angeles, CA 90045
(310) 338-4590
Email: mdillon@lmu.edu